

RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: 10/594,554
Source: IFWP
Date Processed by STIC: 4/13/06

ENTERED



IFWP

RAW SEQUENCE LISTING DATE: 04/13/2006
PATENT APPLICATION: US/10/574,554 **TIME:** 10:16:08

Input Set : A:\Sequence.ST25.txt
Output Set: N:\CRF4\04132006\J574554.raw

```

3 <110> APPLICANT: De Maria, Leonardo
4     Andersen, Carsten
5     Christensen, Lars Lehmann Hylling
6     Lassen, Soren Flensted
7     Ostergaard, Peter Rahbek
9 <120> TITLE OF INVENTION: Protease Variants
11 <130> FILE REFERENCE: 10508.204-US
C--> 13 <140> CURRENT APPLICATION NUMBER: US/10/574,554
C--> 13 <141> CURRENT FILING DATE: 2006-04-03
13 <160> NUMBER OF SEQ ID NOS: 21
15 <170> SOFTWARE: PatentIn version 3.3
17 <210> SEQ ID NO: 1
18 <211> LENGTH: 1596
19 <212> TYPE: DNA
20 <213> ORGANISM: Nocardiopsis sp. NRRL 18262 ("Protease 10")
23 <220> FEATURE:
24 <221> NAME/KEY: CDS
25 <222> LOCATION: (318)..(1463)
27 <220> FEATURE:
28 <221> NAME/KEY: sig_peptide
29 <222> LOCATION: (318)..(404)
31 <220> FEATURE:
32 <221> NAME/KEY: mat_peptide
33 <222> LOCATION: (900)..(1463)
35 <400> SEQUENCE: 1
36 acgttggta cgggtacccgg tgtccgcatt tggccagaat gcccccttgc gacagggAAC 60
38 ggattcggtc ggttagcgcat cgactccgac aaccgcgagg tggccgttcg cgtcgccACG 120
40 ttctgcgacc gtcatgcgac ccatcatcgg gtgaccccac cgagctctga atggtccACC 180
42 gttctgacgg tctttccctc accaaaacgt gcacctatgg ttaggacgtt gtttaccgAA 240
44 tgtctcggtg aacgacaggg gccggacggT attcggcccc gatccccCGT tgatccccCC 300
46 aggagagtag ggaccccc atg cga ccc tcc ccc gtt gtc tcc gcc atc ggt 350
47                               Met Arg Pro Ser Pro Val Val Ser Ala Ile Gly
48                               -190           -185
50 acg gga gcg ctg gcc ttc ggt ctg gcg ctg tcc ggt acc ccg ggt 395
51 Thr Gly Ala Leu Ala Phe Gly Leu Ala Leu Ser Gly Thr Pro Gly
52                               -180           -175           -170
54 gcc ctc gcg gcc acc gga gcg ctc ccc cag tca ccc acc ccg gag 440
55 Ala Leu Ala Ala Thr Gly Ala Leu Pro Gln Ser Pro Thr Pro Glu
56                               -165           -160           -155
58 gcc gac gcg gtc tcc atg cag gag gcg ctc cag cgc gac ctc gac 485
59 Ala Asp Ala Val Ser Met Gln Glu Ala Leu Gln Arg Asp Leu Asp
60                               -150           -145           -140
62 ctg acc tcc gcc gag gcc gag gag ctg ctg gcc gcc cag gac acc 530

```

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/574,554

DATE: 04/13/2006
TIME: 10:16:08

Input Set : A:\Sequence.ST25.txt
Output Set: N:\CRF4\04132006\J574554.raw

63	Leu	Thr	Ser	Ala	Glu	Ala	Glu	Glu	Leu	Leu	Ala	Ala	Gln	Asp	Thr		
64				-135					-130						-125		
66	gcc	tcc	gag	gtc	gac	gag	gcc	gcg	gcc	gag	gcc	gcc	ggg	gac	gcc	575	
67	Ala	Phe	Glu	Val	Asp	Glu	Ala	Ala	Ala	Glu	Ala	Ala	Gly	Asp	Ala		
68				-120					-115						-110		
70	tac	ggc	ggc	tcc	gtc	tcc	gac	acc	gag	agc	ctg	gaa	ctg	acc	gtc	ctg	623
71	Tyr	Gly	Gly	Ser	Val	Phe	Asp	Thr	Glu	Ser	Leu	Glu	Leu	Thr	Val	Leu	
72				-105					-100						-95		
74	gtc	acc	gat	gcc	gcc	gct	gag	gcc	gtg	gag	gcc	acc	ggc	gcc	ggg	671	
75	Val	Thr	Asp	Ala	Ala	Ala	Val	Glu	Ala	Val	Glu	Ala	Thr	Gly	Ala	Gly	
76				-90				-85							-80		
78	acc	gag	ctg	gtc	tcc	tac	ggc	atc	gac	ggt	ctc	gac	gag	atc	gtc	cag	719
79	Thr	Glu	Leu	Val	Ser	Tyr	Gly	Ile	Asp	Gly	Leu	Asp	Glu	Ile	Val	Gln	
80				-75				-70							-65		
82	gag	ctc	aac	gcc	gcc	gac	gcc	gtt	ccc	ggt	gtg	gtc	ggc	tgg	tac	ccg	767
83	Glu	Leu	Asn	Ala	Ala	Asp	Ala	Val	Pro	Gly	Val	Val	Gly	Trp	Tyr	Pro	
84	-60			-55				-50							-45		
86	gac	gtg	gct	ggt	gac	acc	gtc	gtc	ctg	gag	gtc	ctg	gag	ggt	tcc	gga	815
87	Asp	Val	Ala	Gly	Asp	Thr	Val	Val	Leu	Glu	Val	Leu	Glu	Gly	Ser	Gly	
88				-40				-35							-30		
90	gcc	gac	gtc	agc	ggc	ctg	ctc	gct	gac	gcc	gtg	gac	ggc	tgc	gcc	863	
91	Ala	Asp	Val	Ser	Gly	Leu	Leu	Ala	Asp	Ala	Gly	Val	Asp	Ala	Ser	Ala	
92				-25				-20							-15		
94	gtc	gag	gtg	acc	acg	agc	gac	cag	ccc	gag	ctc	tac	gcc	gac	atc	atc	911
95	Val	Glu	Val	Thr	Thr	Ser	Asp	Gln	Pro	Glu	Leu	Tyr	Ala	Asp	Ile	Ile	
96				-10				-5							-1	1	
98	ggt	ggt	ctg	gcc	tac	acc	atg	ggc	ggc	cgc	tgt	tcg	gtc	ggc	tcc	gct	959
99	Gly	Gly	Leu	Ala	Tyr	Thr	Met	Gly	Gly	Arg	Cys	Ser	Val	Gly	Phe	Ala	
100	5			10				15							20		
102	gcc	acc	aac	gcc	gcc	ggt	cag	ccc	ggg	tcc	gtc	acc	gcc	ggt	cac	tgc	1007
103	Ala	Thr	Asn	Ala	Ala	Gly	Gln	Pro	Gly	Phe	Val	Thr	Ala	Gly	His	Cys	
104				25				30							35		
106	ggc	cgc	gtg	ggc	acc	cag	gtg	acc	atc	gac	aac	ggc	agg	ggc	gtc	tcc	1055
107	Gly	Arg	Val	Gly	Thr	Gln	Val	Thr	Ile	Gly	Asn	Gly	Arg	Gly	Val	Phe	
108				40				45							50		
110	gag	cag	tcc	gtc	tcc	ccc	ggc	aac	gac	gct	gcc	tcc	gtc	cgc	ggt	acg	1103
111	Glu	Gln	Ser	Val	Phe	Pro	Gly	Asn	Asp	Ala	Ala	Phe	Val	Arg	Gly	Thr	
112				55				60							65		
114	tcc	aac	tcc	acg	ctg	acc	aac	ctg	gtc	agc	cgc	tac	aac	acc	ggc	ggg	1151
115	Ser	Asn	Phe	Thr	Leu	Thr	Asn	Leu	Val	Ser	Arg	Tyr	Asn	Thr	Gly	Gly	
116				70				75							80		
118	tac	gcc	acg	gtc	gcc	ggt	cac	aac	cag	gcc	ccc	atc	ggc	tcc	tcc	gtc	1199
119	Tyr	Ala	Thr	Val	Ala	Gly	His	Asn	Gln	Ala	Pro	Ile	Gly	Ser	Ser	Val	
120	85			90				95							100		
122	tgc	cgc	tcc	ggc	tcc	acc	acc	ggt	tgg	cac	tgc	ggc	acc	atc	cag	gcc	1247
123	Cys	Arg	Ser	Gly	Ser	Thr	Gly	Trp	His	Cys	Gly	Thr	Ile	Gln	Ala		
124				105				110							115		
126	cgc	ggc	cag	tgc	gtg	agc	tac	ccc	gag	ggc	acc	gtc	acc	aac	atg	acc	1295
127	Arg	Gly	Gln	Ser	Val	Ser	Tyr	Pro	Glu	Gly	Thr	Val	Thr	Asn	Met	Thr	

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/574,554

DATE: 04/13/2006
TIME: 10:16:08

Input Set : A:\Sequence.ST25.txt
Output Set: N:\CRF4\04132006\J574554.raw

128	120	125	130	
130	cgg acc acc	gtg tgc gcc gag ccc ggc gac tcc ggc ggc tcc tac atc		1343
131	Arg Thr Thr Val Cys Ala	Glu Pro Gly Asp Ser Gly Gly Ser Tyr Ile		
132	135	140	145	
134	tcc ggc acc cag gcc cag ggc gtg acc tcc ggc ggc tcc ggc aac tgc			1391
135	Ser Gly Thr Gln Ala Gln Gly Val Thr Ser Gly Gly Ser Gly Asn Cys			
136	150	155	160	
138	cgc acc ggc ggg acc acc ttc tac cag gag gtc acc ccc atg gtg aac			1439
139	Arg Thr Gly Gly Thr Thr Phe Tyr Gln Glu Val Thr Pro Met Val Asn			
140	165	170	175	180
142	tcc tgg ggc gtc cgt ctc cgg acc tgatccccgc ggttccaggc ggaccgacgg			1493
143	Ser Trp Gly Val Arg Leu Arg Thr			
144		185		
146	tcgtgacctg agtaccaggc gtcccccgcg cttccagcgg cgtccgcacc ggggtgggac			1553
148	cggcggtggc cacggccca cccgtgaccg gaccgcccgg cta			1596
151	<210> SEQ ID NO: 2			
152	<211> LENGTH: 382			
153	<212> TYPE: PRT			
154	<213> ORGANISM: Nocardiopsis sp. NRRL 18262 ("Protease 10")			
156	<400> SEQUENCE: 2			
158	Met Arg Pro Ser Pro Val Val Ser Ala Ile Gly Thr Gly Ala Leu			
159		-190	-185	-180
162	Ala Phe Gly Leu Ala Leu Ser Gly Thr Pro Gly Ala Leu Ala Ala			
163		-175	-170	-165
166	Thr Gly Ala Leu Pro Gln Ser Pro Thr Pro Glu Ala Asp Ala Val			
167		-160	-155	-150
170	Ser Met Gln Glu Ala Leu Gln Arg Asp Leu Asp Leu Thr Ser Ala			
171		-145	-140	-135
174	Glu Ala Glu Glu Leu Leu Ala Ala Gln Asp Thr Ala Phe Glu Val			
175		-130	-125	-120
178	Asp Glu Ala Ala Ala Glu Ala Ala Gly Asp Ala Tyr Gly Gly Ser			
179		-115	-110	-105
182	Val Phe Asp Thr Glu Ser Leu Glu Leu Thr Val Leu Val Thr Asp Ala			
183		-100	-95	-90
186	Ala Ala Val Glu Ala Val Glu Ala Thr Gly Ala Gly Thr Glu Leu Val			
187		-85	-80	-75
190	Ser Tyr Gly Ile Asp Gly Leu Asp Glu Ile Val Gln Glu Leu Asn Ala			
191		-70	-65	-60
194	Ala Asp Ala Val Pro Gly Val Val Gly Trp Tyr Pro Asp Val Ala Gly			
195		-55	-50	-45
198	Asp Thr Val Val Leu Glu Val Leu Glu Gly Ser Gly Ala Asp Val Ser			
199		-40	-35	-30
202	-25			
202	Gly Leu Leu Ala Asp Ala Gly Val Asp Ala Ser Ala Val Glu Val Thr			
203		-20	-15	-10
206	Thr Ser Asp Gln Pro Glu Leu Tyr Ala Asp Ile Ile Gly Gly Leu Ala			
207		-5	-1	1
210	Tyr Thr Met Gly Gly Arg Cys Ser Val Gly Phe Ala Ala Thr Asn Ala			
211		10	15	20
214	Ala Gly Gln Pro Gly Phe Val Thr Ala Gly His Cys Gly Arg Val Gly			

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/574,554

DATE: 04/13/2006
TIME: 10:16:08

Input Set : A:\Sequence.ST25.txt
Output Set: N:\CRF4\04132006\J574554.raw

```

215 25          30          35          40
218 Thr Gln Val Thr Ile Gly Asn Gly Arg Gly Val Phe Glu Gln Ser Val
219          45          50          55
222 Phe Pro Gly Asn Asp Ala Ala Phe Val Arg Gly Thr Ser Asn Phe Thr
223          60          65          70
226 Leu Thr Asn Leu Val Ser Arg Tyr Asn Thr Gly Gly Tyr Ala Thr Val
227          75          80          85
230 Ala Gly His Asn Gln Ala Pro Ile Gly Ser Ser Val Cys Arg Ser Gly
231          90          95         100
234 Ser Thr Thr Gly Trp His Cys Gly Thr Ile Gln Ala Arg Gly Gln Ser
235 105          110          115          120
238 Val Ser Tyr Pro Glu Gly Thr Val Thr Asn Met Thr Arg Thr Thr Val
239          125          130          135
242 Cys Ala Glu Pro Gly Asp Ser Gly Gly Ser Tyr Ile Ser Gly Thr Gln
243          140          145          150
246 Ala Gln Gly Val Thr Ser Gly Gly Ser Gly Asn Cys Arg Thr Gly Gly
247          155          160          165
250 Thr Thr Phe Tyr Gln Glu Val Thr Pro Met Val Asn Ser Trp Gly Val
251          170          175          180
254 Arg Leu Arg Thr
255 185
258 <210> SEQ ID NO: 3
259 <211> LENGTH: 1065
260 <212> TYPE: DNA
261 <213> ORGANISM: Nocardiopsis dassonvillei subspecies dassonvillei DSM 43235
("Protease 18")
264 <220> FEATURE:
265 <221> NAME/KEY: CDS
266 <222> LOCATION: (1)..(1062)
268 <220> FEATURE:
269 <221> NAME/KEY: mat_peptide
270 <222> LOCATION: (499)..(1062)
272 <400> SEQUENCE: 3
273 gct ccg gcc ccc gtc ccc cag acc ccc gtc gcc gac gac agc gcc      45
274 Ala Pro Ala Pro Val Pro Gln Thr Pro Val Ala Asp Asp Ser Ala
275          -165          -160          -155
277 gcc agc atg acc gag gcg ctc aag cgc gac ctc gac ctc acc tcg      90
278 Ala Ser Met Thr Glu Ala Leu Lys Arg Asp Leu Asp Leu Thr Ser
279          -150          -145          -140
281 gcc gag gcc gag gag ctt ctc tcg gcg cag gaa gcc gcc atc gag      135
282 Ala Glu Ala Glu Glu Leu Leu Ser Ala Gln Glu Ala Ala Ile Glu
283          -135          -130          -125
285 acc gac gcc gag gcc acc gag gcc gcg ggc gag gcc tac ggc ggc      180
286 Thr Asp Ala Glu Ala Thr Glu Ala Ala Gly Glu Ala Tyr Gly Gly
287          -120          -115          -110
289 tca ctg ttc gac acc gag acc ctc gaa ctc acc gtg ctg gtc acc gac      228
290 Ser Leu Phe Asp Thr Glu Thr Leu Glu Leu Thr Val Leu Val Thr Asp
291          -105          -100          -95
293 gcc tcc gcc gtc gag gcg gtc gag gcc acc gga gcc cag gcc acc gtc      276
294 Ala Ser Ala Val Glu Ala Val Glu Ala Thr Gly Ala Gln Ala Thr Val

```

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/574,554

DATE: 04/13/2006
TIME: 10:16:08

Input Set : A:\Sequence.ST25.txt
Output Set: N:\CRF4\04132006\J574554.raw

295	-90	-85	-80	-75	
297	gtc tcc cac ggc acc gag ggc ctg acc gag gtc gtg gag gac ctc aac				324
298	Val Ser His Gly Thr Glu Gly Leu Thr Glu Val Val Glu Asp Leu Asn				
299	-70	-65	-60		
301	ggc gcc gag gtt ccc gag agc gtc ctc ggc tgg tac ccg gac gtg gag				372
302	Gly Ala Glu Val Pro Glu Ser Val Leu Gly Trp Tyr Pro Asp Val Glu				
303	-55	-50	-45		
305	agc gac acc gtc gtg gtc gag gtg ctg gag ggc tcc gac gcc gac gtc				420
306	Ser Asp Thr Val Val Glu Val Leu Glu Gly Ser Asp Ala Asp Val				
307	-40	-35	-30		
309	gcc gcc ctg ctc gcc gac gcc ggt gtg gac tcc tcc tcg gtc cgg gtg				468
310	Ala Ala Leu Leu Ala Asp Ala Gly Val Asp Ser Ser Val Arg Val				
311	-25	-20	-15		
313	gag gag gcc gag gag gcc ccg cag gtc tac gcc gac atc atc ggc ggc				516
314	Glu Glu Ala Glu Ala Pro Gln Val Tyr Ala Asp Ile Ile Gly Gly				
315	-10	-5	-1 1	5	
317	ctg gcc tac tac atg ggc ggc cgc tgc tcc gtc ggc ttc gcc gcg acc				564
318	Leu Ala Tyr Tyr Met Gly Gly Arg Cys Ser Val Gly Phe Ala Ala Thr				
319	10	15	20		
321	aac agc gcc ggt cag ccc ggt ttc gtc acc gcc ggc cac tgc ggc acc				612
322	Asn Ser Ala Gly Gln Pro Gly Phe Val Thr Ala Gly His Cys Gly Thr				
323	25	30	35		
325	gtc ggc acc ggc gtg acc atc ggc aac ggc acc ggc acc ttc cag aac				660
326	Val Gly Thr Gly Val Thr Ile Gly Asn Gly Thr Gly Thr Phe Gln Asn				
327	40	45	50		
329	tcg gtc ttc ccc ggc aac gac gcc gcc ttc gtc cgc ggc acc tcc aac				708
330	Ser Val Phe Pro Gly Asn Asp Ala Ala Phe Val Arg Gly Thr Ser Asn				
331	55	60	65	70	
333	tcc acc ctg acc aac ctg gtc tcg cgc tac aac tcc ggc ggc tac cag				756
334	Phe Thr Leu Thr Asn Leu Val Ser Arg Tyr Asn Ser Gly Gly Tyr Gln				
335	75	80	85		
337	tcg gtg acc ggt acc agc cag gcc ccg ggc tcc gtc gcc gtg tgc cgc				804
338	Ser Val Thr Gly Thr Ser Gln Ala Pro Ala Gly Ser Ala Val Cys Arg				
339	90	95	100		
341	tcc ggc tcc acc acc ggc tgg cac tgc ggc acc atc cag gcc cgc aac				852
342	Ser Gly Ser Thr Thr Gly Trp His Cys Gly Thr Ile Gln Ala Arg Asn				
343	105	110	115		
345	cag acc gtg cgc tac ccg cag ggc acc gtc tac tcg ctc acc cgc acc				900
346	Gln Thr Val Arg Tyr Pro Gln Gly Thr Val Tyr Ser Leu Thr Arg Thr				
347	120	125	130		
349	aac gtg tgc gcc gag ccc ggc gac tcc ggc ggt tcg ttc atc tcc ggc				948
350	Asn Val Cys Ala Glu Pro Gly Asp Ser Gly Gly Ser Phe Ile Ser Gly				
351	135	140	145	150	
353	tcg cag gcc cag ggc gtc acc tcc ggc ggc tcc ggc aac tgc tcc gtc				996
354	Ser Gln Ala Gln Gly Val Thr Ser Gly Gly Ser Gly Asn Cys Ser Val				
355	155	160	165		
357	ggc ggc acg acc tac tac cag gag gtc acc ccg atg atc aac tcc tgg				1044
358	Gly Gly Thr Thr Tyr Tyr Gln Glu Val Thr Pro Met Ile Asn Ser Trp				
359	170	175	180		

RAW SEQUENCE LISTING ERROR SUMMARY
PATENT APPLICATION: US/10/574,554

DATE: 04/13/2006
TIME: 10:16:09

Input Set : A:\Sequence.ST25.txt
Output Set: N:\CRF4\04132006\J574554.raw

Invalid Line Length:

The rules require that a line not exceed 72 characters in length. This includes spaces.

Seq#:3; Line(s) 261
Seq#:4; Line(s) 369

Invalid <213> Response:

Use of "Artificial" only as "<213> Organism" response is incomplete,
per 1.823(b) of New Sequence Rules. Valid response is Artificial Sequence.

Seq#:11,12,13,14,15,16,17,18,19,20,21

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/574,554

DATE: 04/13/2006

TIME: 10:16:09

Input Set : A:\Sequence.ST25.txt

Output Set: N:\CRF4\04132006\J574554.raw

L:13 M:270 C: Current Application Number differs, Replaced Current Application No
L:13 M:271 C: Current Filing Date differs, Replaced Current Filing Date